



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/844,040	04/27/2001	Brian Seong-Gon Kim	089984-0276267	2687

27271 7590 06/26/2003
PILLSBURY WINTHROP LLP
DOCKET/CALENDAR DEPARTMENT
50 FREMONT STREET
SAN FRANCISCO, CA 94105-2230

[REDACTED] EXAMINER

CHEN, CHONGSHAN

[REDACTED] ART UNIT

[REDACTED] PAPER NUMBER

2172

DATE MAILED: 06/26/2003

4

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application 09/844,040	Applicant(s) KIM ET AL.
	Examiner Chongshan Chen	Art Unit 2172

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on _____.
- 2a) This action is **FINAL**. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-39 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 1-39 is/are rejected.
- 7) Claim(s) _____ is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on _____ is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) The proposed drawing correction filed on _____ is: a) approved b) disapproved by the Examiner.
If approved, corrected drawings are required in reply to this Office action.
- 12) The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

- 13) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
 - a) All
 - b) Some *
 - c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.
- 14) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
 - a) The translation of the foreign language provisional application has been received.
- 15) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

- 1) Notice of References Cited (PTO-892)
- 2) Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) Information Disclosure Statement(s) (PTO-1449) Paper No(s) 2.
- 4) Interview Summary (PTO-413) Paper No(s) _____.
- 5) Notice of Informal Patent Application (PTO-152)
- 6) Other: _____

DETAILED ACTION

1. Claims 1-39 are pending in this Office Action.

Information Disclosure Statement

2. The reference cited in the IDS, PTO-1449, Paper No. 2, has been considered.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1-39 are rejected under 35 U.S.C. 103(a) as being unpatentable over Krellenstein (5,924,090) in view of Zamir et al. (“Zamir”, “Web Document Clustering: A Feasibility Demonstration”).

As per claim 1, Krellenstein discloses a method of categorizing an initial collection of documents, each document being represented by a string of characters, the method comprising the steps of:

identifying predefined characters in the string of characters from the documents in the initial collection of documents to form identified characters (Krellenstein, Fig. 2);

constructing a number of categories from the preprocessed collection of documents (Krellenstein, Fig. 2, col. 2, lines 56-65); and

assigning each document in the preprocessed collection of documents to a category to form a hierarchy of categories of documents (Krellenstein, Fig. 2, col. 2, lines 56-65).

Krellenstein does not explicitly disclose changing the identified characters in the documents in the initial collection of documents to form a preprocessed collection of documents. Zamir discloses changing the identified characters in the documents in the initial collection of documents to form a preprocessed collection of documents (Zamir, page 3, 3.1 Step 1 – Document “Cleaning”). Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to combine Zamir with Krellenstein in order to identify key phrase.

As per claim 2, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 1, and further disclose

clearing a temporary category and selecting a seed document as a first document of the temporary category (Zamir, page 3, 3.2 Step 2 – Identifying Base Clusters);

collecting documents from the preprocessed collection of documents that are similar to the seed document into the temporary category (Krellenstein, Fig. 2);

testing to determine if there are enough documents in the temporary category to merit construction of a new category (Krellenstein, Fig. 2);

constructing the new category and generating a heading for the new category if there are enough documents in the temporary category to merit construction (Krellenstein, Fig. 2);

assigning the seed document to a category reserved for documents not belonging to any specific category if there are not enough documents in the temporary category (Krellenstein, Fig. 2); and

marking the documents assigned to any category in the preprocessed collection of documents as processed (Krellenstein, Fig. 2).

As per claim 3, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose the predefined characters include punctuation marks, and the changing step removes the punctuation marks from the string of characters (Zamir, page 3, 3.1 Step 1 – Document “Cleaning”).

As per claim 4, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, except for explicitly disclosing the predefined characters include upper-case characters, and the changing step replaces upper-case characters with lower-case characters. However, Zamir discloses changing the identified characters by reducing plural to singular and deleting numbers and punctuations (Zamir, page 3, 3.1 Step 1 – Document “Cleaning”). It would have been obvious to one of ordinary skill in the art at the time the invention was made to replaces upper-case characters with lower-case characters in order to identify key phrase and enhance user readability. Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to replaces upper-case characters with lower-case characters in order to identify key phrase and enhance user readability.

As per claim 5, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose the predefined characters include non-root words, and the changing step replaces the non-root words with root words (Zamir, page 3, 3.1 Step 1 – Document “Cleaning”).

As per claim 6, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, except for explicitly disclosing the predefined characters include abbreviations, and the changing step replaces the abbreviations with original words. However, Zamir discloses changing the identified characters by reducing plural to singular and deleting numbers and

punctuations (Zamir, page 3, 3.1 Step 1 – Document “Cleaning”). It would have been obvious to one of ordinary skill in the art at the time the invention was made to replace the abbreviations with original words in order to identify key phrase and enhance user readability. Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to replace the abbreviations with original words in order to identify key phrase and enhance user readability.

As per claim 7, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose except for explicitly disclosing the predefined characters include articles, and the changing step removes the articles from the string of characters (Zamir, page 3, 3.1 Step 1 – Document “Cleaning”).

As per claim 8, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose the collecting step further includes the step of loading a character string from the seed document into a memory location to initialize the values of a number of category properties for the temporary category (Zamir, page 3, 3.2 Step 2 – Identifying Base Clusters).

As per claim 9, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 8, and further disclose

determining if there are documents in the preprocessed collection of documents that have not been processed with respect to the temporary category (Krellenstein, Fig. 2);

if there are documents in the preprocessed collection of documents that have not been processed with respect to the temporary category, selecting a next document from the

preprocessed collection of documents and measuring a similarity with a similarity test between the selected document and a number of current category properties (Krellenstein, Fig. 2);

including the selected document in the temporary category if the selected document passes the similarity test (Zamir, page 4, 3.3 Step 3 – Combining Base Clusters);
updating the values of the number of category properties of the temporary category when the selected document is included (Krellenstein, Fig. 2); and

rejecting the selected document if the selected document fails the similarity test (Zamir, page 4, 3.3 Step 3 – Combining Base Clusters).

As per claim 10, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 9, and further disclose repeating the steps of claim 9 for all documents in preprocessed collection of documents (Krellenstein, Fig. 2).

As per claim 11, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose collecting more similar documents from a number of existing categories (Krellenstein, Fig. 2).

As per claim 12, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 11, and further disclose

determining if there are more documents in a number of existing categories that have not been processed with respect to the temporary category (Krellenstein, Fig. 2);

if there are documents in the number of existing categories that have not been processed with respect to the temporary category, selecting a next document from the number of existing categories as a selected document and measuring a similarity with a similarity test between the

selected document and a number of current category properties (Krellenstein, Fig. 2, Zamir, page 4, 3.3 Step 3 – Combining Base Clusters);

including the selected document in the temporary category if the selected document passes the similarity test (Krellenstein, Fig. 2, Zamir, page 4, 3.3 Step 3 – Combining Base Clusters); and

rejecting the selected document if the selected document fails the similarity test (Krellenstein, Fig. 2, Zamir, page 4, 3.3 Step 3 – Combining Base Clusters).

As per claim 13, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 12, and further disclose repeating the steps of claim 12 for all documents in the number of existing categories (Krellenstein, Fig. 2).

As per claim 14, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 8, and further disclose the category properties includes a string of characters selected from the group consisting of a longest common substring in the title, a longest common substring in the body; and a document type index measured as list of fractional numbers for each document type (Zamir, page 3, 3.2, Step 2 – Identifying Base Clusters).

As per claim 15, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 14, and further disclose categorizing documents into categories (Krellenstein, Fig. 2), the documents inherently includes news article, technical documents, and poems.

As per claim 16, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose making sub-categories if there are too many documents in a given category; and post-processing the number of categorized lists of documents (Krellenstein, col. 5, lines 30-41).

As per claim 17, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 16, and further disclose merging two categories that each have a heading where there is too much overlap in the headings of the two categories; and promoting sub-categories to an upper level in a hierarchy when there are not enough categories in the upper level (Krellenstein, col. line 66 – col. 7, line 27).

As per claim 18, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose the seed document is a first document in the preprocessed collection of documents (Zamir, page 3, 3.2 Step 2 – Identifying Base Clusters).

As per claim 19, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose the seed document is a document with a highest rank value among the documents not marked as processed in the preprocessed collection of documents (Krellenstein, col. 8, lines 16-28).

As per claim 20, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, and further disclose the temporary category is tested to determine if there are enough documents in the temporary category to merit construction of a new category by accumulating the weight of each document when each document can contribute uniform weight or different weight based on the rank value of each document with higher ranked document given more weight (Krellenstein, Fig. 2).

As per claim 21, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 2, except for explicitly disclosing the heading is a longest common substring in a title. Zamir disclose using common phrase to cluster document (Zamir, page 3, 3.2 Step 2 – Identifying Base Clusters). It would have been obvious to one of ordinary skill in the art at the

time the invention was made to use the longest common substring in a title as category heading because the longest common phrase in the title describes the topic of the category. Therefore, it would have been obvious to one of ordinary skill in the art at the time the invention was made to use the longest common substring in a title as category heading because the longest common phrase in the title describes the topic of the category.

Claim 22 is rejected on grounds corresponding to the reasons given above for claim 21.

As per claim 23, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 1, and further disclose determining if an anchor-text character string is available for the documents in the initial collection of documents; and attaching an anchor-text character string to the string of characters that represents the documents in the initial collection of documents when the anchor text character string is available (Zamir, page 4, 3.3 Step 3 – Combing Base Clusters).

As per claim 24, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 23, and further disclose the anchor-text character string is a text used most frequently by hypertext documents (Zamir, page 4, 3.3 Step 3 – Combing Base Clusters).

As per claim 25, Krellenstein and Zamir teach all the claimed subject matters as discussed in claim 23, and further disclose the anchor-text character string is a text with a highest partial extrinsic rank value (Zamir, page 4, 3.3 Step 3 – Combing Base Clusters).

Claims 26-39 are rejected on grounds corresponding to the reasons given above for claims 1-25.

Conclusion

5. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Chongshan Chen whose telephone number is 703-305-8319. The examiner can normally be reached on Mon. - Fri. 8:00 - 4:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Kim Vu can be reached on 703-305-4393. The fax phone numbers for the organization where this application or proceeding is assigned are 703-746-7239 for regular communications and 703-746-7238 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

CC
June 13, 2003

Shahid Al Alam
SHAHID AL ALAM
PATENT EXAMINER